The Report Committee for Tian Lan
Certifies that this is the approved version of the following report:

Analysis of circular data in the dynamic model and mixture of von Mises distributions

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Carlos M. Carvalho

____
Timothy H. Keitt

Analysis of circular data in the dynamic model and mixture of von Mises distributions

by

Tian Lan, B.S.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2013

Acknowledgements

When composing this report, I have obtained assistance and support from many people, without whom this report may not have been completed. First of all, I wish to express my sincere gratitude to Dr. Carlos M. Carvalho from the McCombs School of Business for the endless advice and guidance for me at every step along the way. I am grateful for his patient teaching and careful revising of the drafts of this report. Then I would like to thank Dr. Timothy H. Keitt for the initial motivation for the idea and topic of this report.

Abstract

Analysis of circular data in the dynamic model and mixture of von Mises distributions

Tian Lan, M.S. Stat.

The University of Texas at Austin, 2013

Supervisor: Carlos M. Carvalho

Analysis of circular data becomes more and more popular in many fields of studies. In this report, I present two statistical analysis of circular data using von Mises distributions. Firstly, the maximization-expectation algorithm is reviewed and used to classify and estimate circular data from the mixture of von Mises distributions. Secondly, Forward Filtering Backward Smoothing method via particle filtering is reviewed and implemented when circular data appears in the dynamic state-space models.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Data that deal with or are measured in the form of axes (lines through the origin), rotations, or directions (unit vectors) appear in the Science everywhere. They are commonly found in Biology, Geography, Geology, Medicine, Oceanography, and in many other fields (Fisher 1993). For example, to investigate the occurrence of reverse autumn migration among 20 passerine bird species from a coastal site in southwesternmost Sweden, Akesson et al. (1996) measured the mean angle of the direction of birds' recoveries and recorded as circular data. Capaccioni et al. (1997) also analyzed circular data in their study. Particularly, using image analysis applied to textural investigation on sedimentary or volcanic rocks, Capaccioni et al. obtained fabric data and calculated the theoretical circumference and the best-fit ellipse, as well as all the geometrical parameters. Due to the massive usage of circular data in different fields of research, there has been a vigorous development of statistical methodologies in analyzing and modeling circular data.

THE VON MISES DISTRIBUTION

Circular data have some special properties when compared with regular data. For example, if we have two data points measured in degrees 1° and 359°, then the average direction of these two points should be 0° or 360°. However, the arithmetic mean, 180°, can give misleading conclusions. As a result, distinctive probability models specified by suitable estimation of parameters are used to quantify circular data. The von Mises distribution is one of the most common models for samples of circular data. It is a continuous probability distribution on the circle. Symmetric and unimodal, as the analogue on the circle of the Normal distribution on the real line, the von Mises distribution is a close approximation to the Wrapped Normal distribution (Fisher 1993).

The von Mises probability density function for the angle θ is given by:

$f(\theta) = [2\pi I_0(\lambda)]^{-1}\exp[\lambda\cos(\theta-\mu)],\ 0 \le \theta < 2\pi\ ,\ 0 \le \lambda < \infty$

where

$$I_0(\lambda) = (2\pi)^{-1}\int_0^{2\pi}\exp[\lambda\cos(\emptyset-\mu)]d\emptyset$$

is the modified Bessel function of order zero.

The parameters μ and $1/\lambda$ are analogous to the mean and variance in the Normal distribution. In particular, μ is a measure of mean direction, where the distribution is clustered around. $1/\lambda$ describes the circular dispersion. As $\lambda \rightarrow 0$, the distribution converges to the uniform distribution; however, as $\lambda \rightarrow \infty$, the distribution tends to the point distribution concentrated in the direction μ.

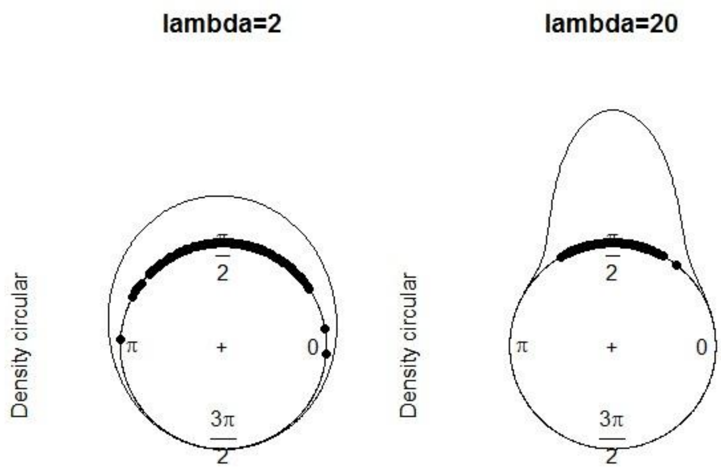

Figure 1.1: Probability density functions of von Mises distributions with the same mean but different concentration parameters.

There is no simple closed form for the cumulative density function:

$$F(\theta) = [2\pi I_0(\lambda)]^{-1}\int_0^{\theta} exp[\lambda cos(\emptyset-\mu)]\,d\emptyset$$

MAXIMUM LIKELIHOOD ESTIMATION

Extensive research has been done to estimate the parameters for the von Mises distribution. The usual maximum likelihood estimator (MLE) $\hat{\mu}$, the mean direction μ, is the sample mean direction $\bar{\theta}$. To get the sample mean direction, $\bar{\theta}$, first calculate:

$$C = \sum_{i=1}^{n} cos\theta_i,\ S = \sum_{i=1}^{n} sin\theta_i, \text{ and } R^2 = C^2 + S^2.$$

Then the direction $\bar{\theta}$ of $\theta_1, \dots, \theta_n$ is given by

$$cos\bar{\theta} = C/R,\ sin\bar{\theta} = S/R.$$

The maximum likelihood estimator $\hat{\lambda}_{MLE}$ of λ is the solution of the equation

$$A_1(\hat{\lambda}_{MLE}) = R/n,$$

where $A_1(\hat{\lambda}_{MLE}) = I_1(\hat{\lambda}_{MLE})/I_0(\hat{\lambda}_{MLE})$ is the ratio of two modified Bessel functions (Fisher 1993).

In this report, I present some statistical analysis of circular data. In particular, Chapter 1 talks about fitting circular data with a finite number of von Mises distributions through Expectation-Maximization algorithm. Chapter 2 talks about the analysis of circular data in the dynamic models through forward Filtering Backward Smoothing method.

Chapter 2: Mixtures of the von Mises Distributions Estimation

INTRODUCTION

In this chapter I present a method to estimate a finite number of mixtures of the von Mises distributions. First I review the algorithm and then use a simulated data set to assess the performance of the algorithm. At last, I apply this method to a real data set, in which records patients arrival time on a 24-hour clock.

Consider a set of n independent random circular observations $y_1, \cdots, y_n$ following a K-component mixture of the von Mises Distributions. Specifically,

$$f(y_i) = \sum_{j=1}^{K} p_j f_j(y_i | \mu_j, \lambda_j), i = 1, \cdots, n,$$

where $\sum_{j=1}^{K} p_j = 1$. Our purpose is to estimate all the parameters of each component $\mu = (\mu_1, \cdots, \mu_K)$, $\lambda = (\lambda_1, \cdots, \lambda_K)$, as well as the weights $p = (p_1, \cdots, p_K)$. An example of data and the probability density function is shown in Figure 2.1. In this example, 500 observations are generated from two mixtures of von Mises distributions with parameters $\mu = \left(\frac{\pi}{2}, \frac{7\pi}{5}\right)$, $\lambda = (3, 5)$, and the weights $p = (0.3, 0.7)$.

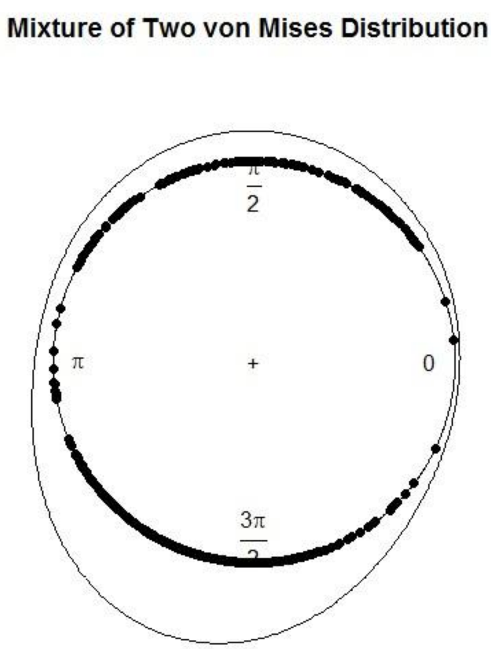

Figure 2.1: Plot of 500 observations generated from a mixture of two von Mises distributions

EXPECTATION-MAXIMIZATION ALGORITHM ESTIMATION

Because the mixture model depends on unobserved latent variables, which assign a given sample to a given component of the mixture, the Expectation-Maximization (EM) algorithm is often used to find the maximum likelihood estimates of the mixture parameters. Calderara et al. (2011) derived the EM steps for a mixture of von Mises distribution and the steps are summarized below:

Expectation Step

Given a mixture of von Mises with K components, the latent variable γ_{ik} of component k for sample y_i can be estimated using the parameter values of the previous iteration (randomly initialized for the first iteration) as follows:

$$\gamma_{ik} = \frac{p_k vM(y_i|\mu_k, \lambda_k)}{\sum_{j=1}^{K} p_j vM\left(y_i|\mu_j, \lambda_j\right)}$$

Maximization Step

Based on the latent variables found on the E step, the M step computes parameters maximizing the expected log-likelihood. The updated new values of the mixture's parameters, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_K)$, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_K)$, and $\boldsymbol{p} = (p_1, \cdots, p_K)$, for each component k, can be estimated as:

$$p_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \gamma_{ik}$$

$$\mu_k = arctan\left(\frac{\sum_{i=1}^{n_k} \gamma_{ik} sin y_i}{\sum_{i=1}^{n_k} \gamma_{ik} cos y_i}\right)$$

$$A(\lambda_k) = \frac{I_1(\lambda_k)}{I_0(\lambda_k)} = \frac{\sum_{i=1}^{n_k} \gamma_{ik} cos(y_i - \mu_k)}{\sum_{i=1}^{n_k} \gamma_{ik}}$$

where n_k is the number of samples in the k^{th} component of the mixture.

The E step and M step are iterated until the convergence when the likelihood does not change too much between two consecutive iterations or until a certain number of iterations is reached.

EM Algorithm

Step 0 (Initiation). Choose $\boldsymbol{\mu}_0$, $\boldsymbol{\lambda}_0$, and $\boldsymbol{p}_0$ arbitrarily.

Step 1 (Expectation). Calculate the latent variable γ_{ik} of component k for each sample $\mathrm{y_i}$.

Step 2 (Maximization). Calculate p_k, μ_k, and λ_k for each component.

EXAMPLE WITH SIMULATED DATA

Consider the example I mentioned in Figure 2.1. 500 observations are generated from two mixtures of von Mises distributions with parameters $\mu = \left(\frac{\pi}{2}, \frac{7\pi}{5}\right)$, $\lambda = (3, 5)$, and the weights $p = (0.3, 0.7)$.

Firstly I fit the data with one von Mises distribution and used MLE to estimate the parameters. Table 2.1 summarizes the estimations. The data is estimated to be centered at 4.1772 or 239.34 degrees. The concentration parameter is 0.8914, indicating that data are very sparse. Figure 2.2 shows the comparison between empirical density and estimated density. In order to perform the goodness of fit test, 500 data are simulated from the von Mises distribution with the estimated parameters shown in Table 2.1. The simulated data are compared with the observed data using Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic is obtained to be 0.198 and the p-value is 6.137×10^{-9}, which is smaller than 0.05. Therefore, I reject the null hypothesis and conclude that two sets of data are not from the same distribution. So MLE does not give an accurate fit to the data.

	μ	**λ**
Estimation	4.17772	0.8914

Table 2.1: The summary of estimations using MLE for simulated data

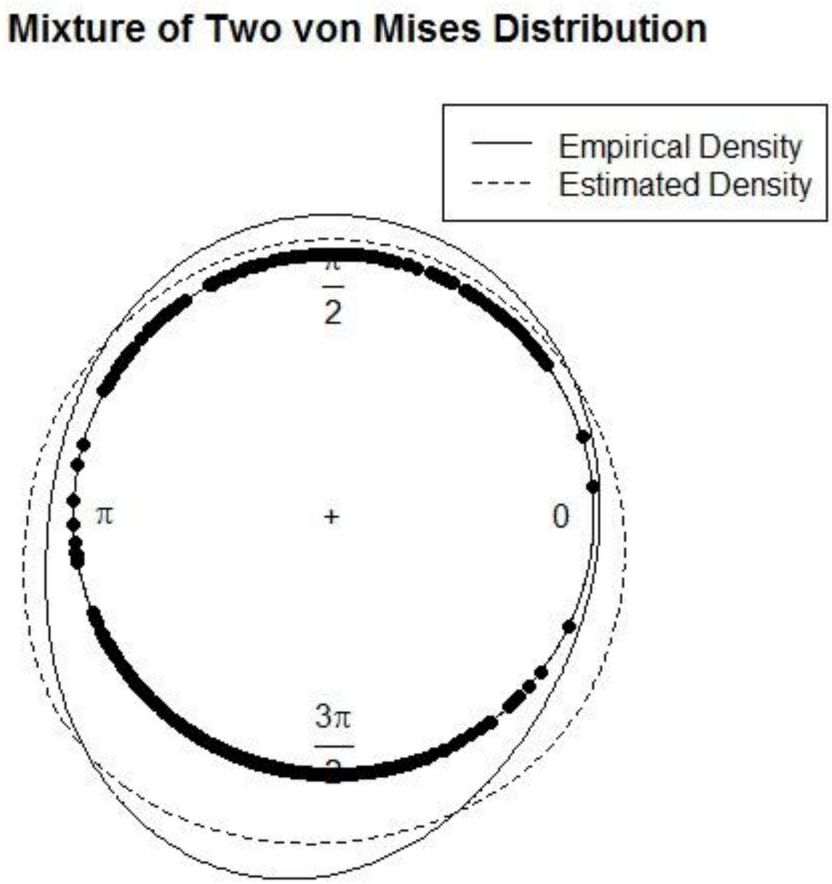

Figure 2.2: Comparison of the empirical density and the estimated density from MLE

Secondly, the EM algorithm is used to estimate the data. The estimations are shown in Table 2.2. In this simulated example, estimations of all the parameters are quite close. Figure 2.3 shows the log likelihood increases over the iterations and reaches the maximum at the 7^{th} iteration, indicating the convergence. Figure 2.4 shows the comparison between empirical density and estimated density using the EM algorithm. As shown in the figure, two densities are very close to each other, indicating a good fit.

In order to perform the goodness of fit test, 500 data are simulated from the mixture of two von Mises distributions with the estimated parameters shown in Table 2.2. The simulated data are compared with the observed data using Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic is obtained to be 0.044 and the p-value is 0.7184. Therefore, I fail to reject the null hypothesis and conclude that two sets of data are from the same distribution. So the EM algorithm gives good estimates fitting the data well.

Parameters	**μ_1**	**μ_2**	**λ_1**	**λ_2**	**p_1**	**p_2**
Estimations	1.555743	4.37453	3.07	5.39	0.2994798	0.7005202
True	1.570796	4.39823	3	5	0.3	0.7

Table 2.2: The comparison between the estimations using EM algorithm and the true parameters

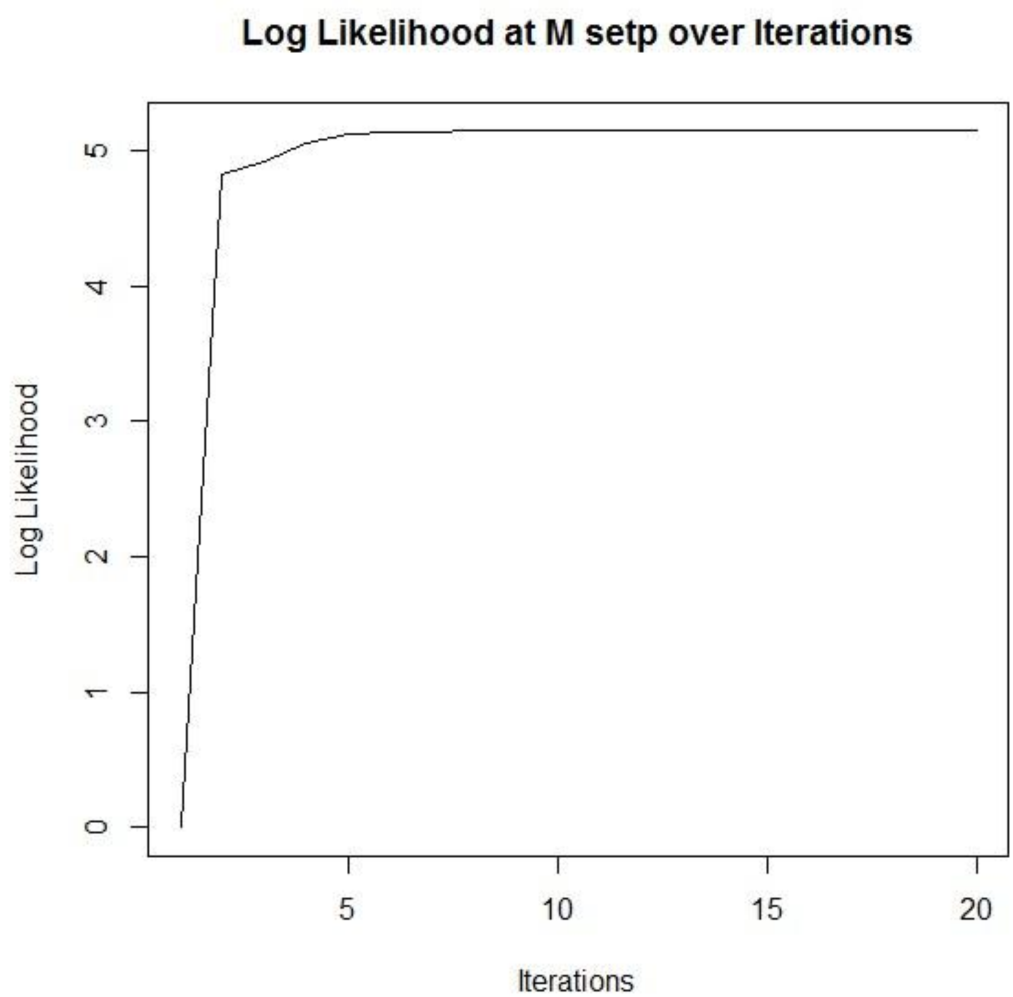

Figure 2.3: The change of log likelihood at M step over iterations.

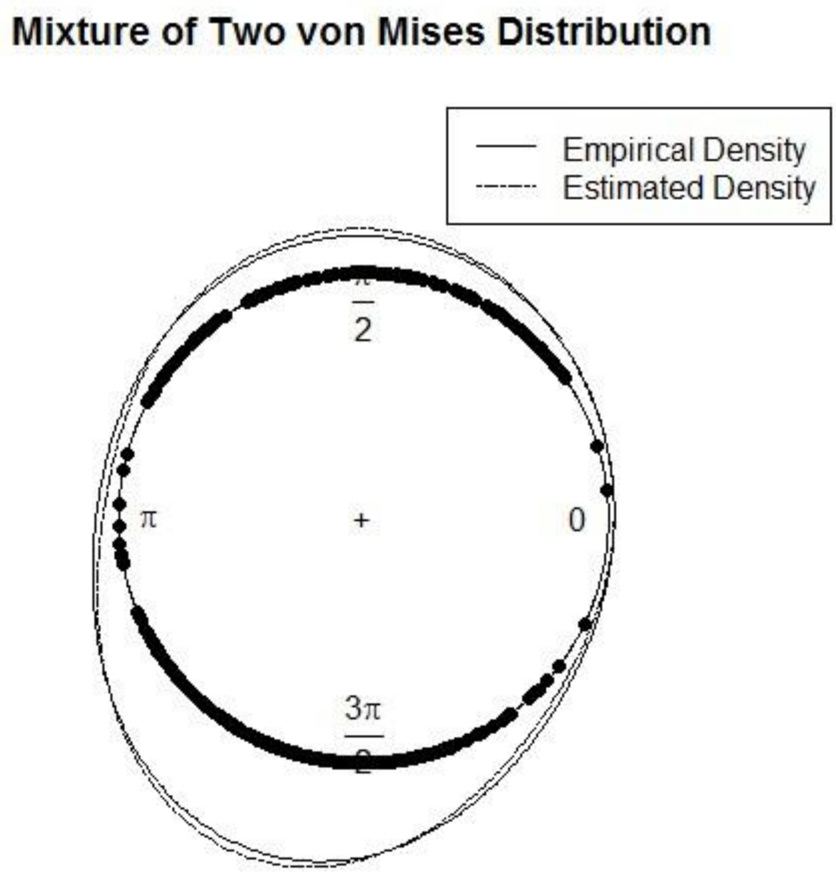

Figure 2.4: Comparison of the empirical density and the estimated density from EM algorithm

PATIENTS ARRIVAL TIMES DATA ANALYSIS

In this section, patients arrival times data is analyzed, which is obtained from Fisher (1993). The data record the arrival times on a 24-hour clock of 254 patients at an intensive care unit, over a period of about 12 months. The arrival times may be regarded as circular measurements, with an arrival time of m minutes after midnight corresponding to a circular measurement of $360 \times m/(24 \times 60)$ degrees; thus 1 degree corresponds to 4 minutes of time. Figure 2.5 shows the plot of patients arrival times data on a unit circle from 0 to 2π, where midnight (12:00 a.m.) is mapped to 0 on the circle. The plot suggests a fairly steady stream between 10:00 a.m. and 10:00 p.m., with some smaller clusters of arrivals about 2:00 a.m. and 5:00 a.m.. The goal is to use finite number of mixture of von Mises distributions to fit this data set.

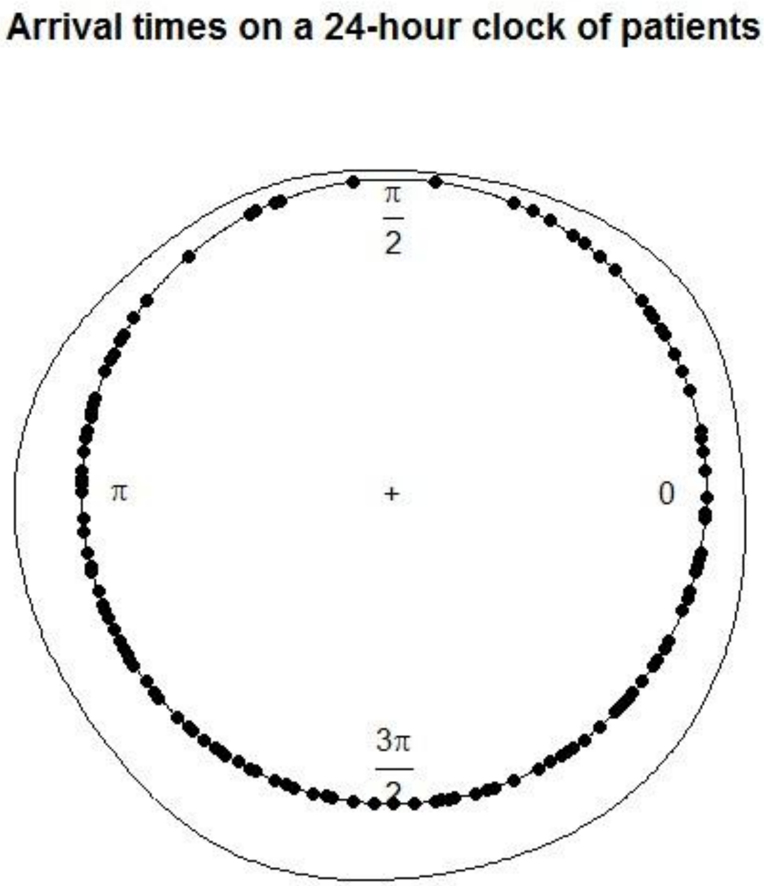

Figure 2.5: Arrival times on a 24-hour clock of 254 patients at an intensive care unit, over a period of about 12 months

Firstly I fit the data set with one von Mises distribution and use MLE to estimate parameters. Table 2.3 summarizes the estimations. The data is estimated to be centered at 4.5182, which is about 17:15 pm.. The concentration parameter is 0.6817. Figure 2.6 shows the comparison between empirical density and estimated density.

In order to perform the goodness of fit test, 500 data are simulated from the von Mises distribution with the estimated parameters shown in Table 2.3. The simulated data are compared with the observed data using Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic is obtained to be 0.0488 and the p-value is 0.8179, which is greater than 0.05. Therefore, I fail to reject the null hypothesis and conclude that two sets of data are from the same distribution. So fitting data with one von Mises distribution and using MLE gives a good fit to the data.

	μ	λ
Estimation	4.5182	0.6817

Table 2.3: The summary of estimations using MLE for patient arrival data

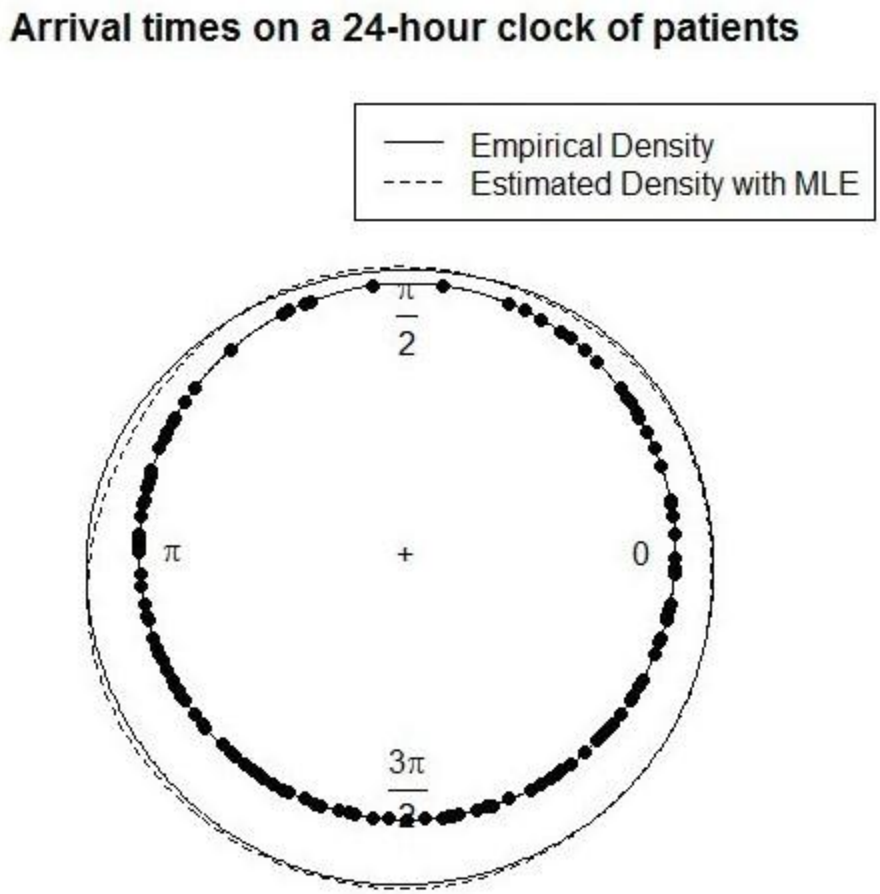

Figure 2.6: Comparison of the empirical density from the data and the estimated density from the von Mises distributions using MLE

Secondly, I fit the data set with a mixture of two von Mises distributions and use EM algorithm to estimate all the parameters. Table 2.4 summarizes the estimations. Figure 2.7 shows the log likelihood increases over the iterations and reaches the maximum at the 3rd iteration, indicating the convergence. The model suggests that the first cluster is at 3.5841, which is about 1:40 p.m.. The second cluster is at 5.472, which is about 9:50 p.m. Both clusters are relatively sparse, where the concentration parameters are 1.87 and 1.05, respectively. However, the second cluster is larger than the first one, where the proportion is 58.94% and 41.06%, respectively. Figure 2.8 shows the comparison of the empirical density from the data and the estimated density from the mixture of two von Mises distributions.

In order to perform the goodness of fit test, 500 data are simulated from the von Mises distribution with the estimated parameters shown in Table 2.4. The simulated data are compared with the observed data using Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic is obtained to be 0.0736 and the p-value is 0.3208, which is greater than 0.05. Therefore, I fail to reject the null hypothesis and conclude that two sets of data are from the same distribution. So fitting data with the mixture of two von Mises distributions gives an good fit to the data.

	μ_1	μ_2	λ_1	λ_2	p_1	p_2
Estimations	3.584056	5.471985	1.87	1.05	0.41057	0.58943

Table 2.4: The summary of estimated parameters of a mixture of two von Mises distributions

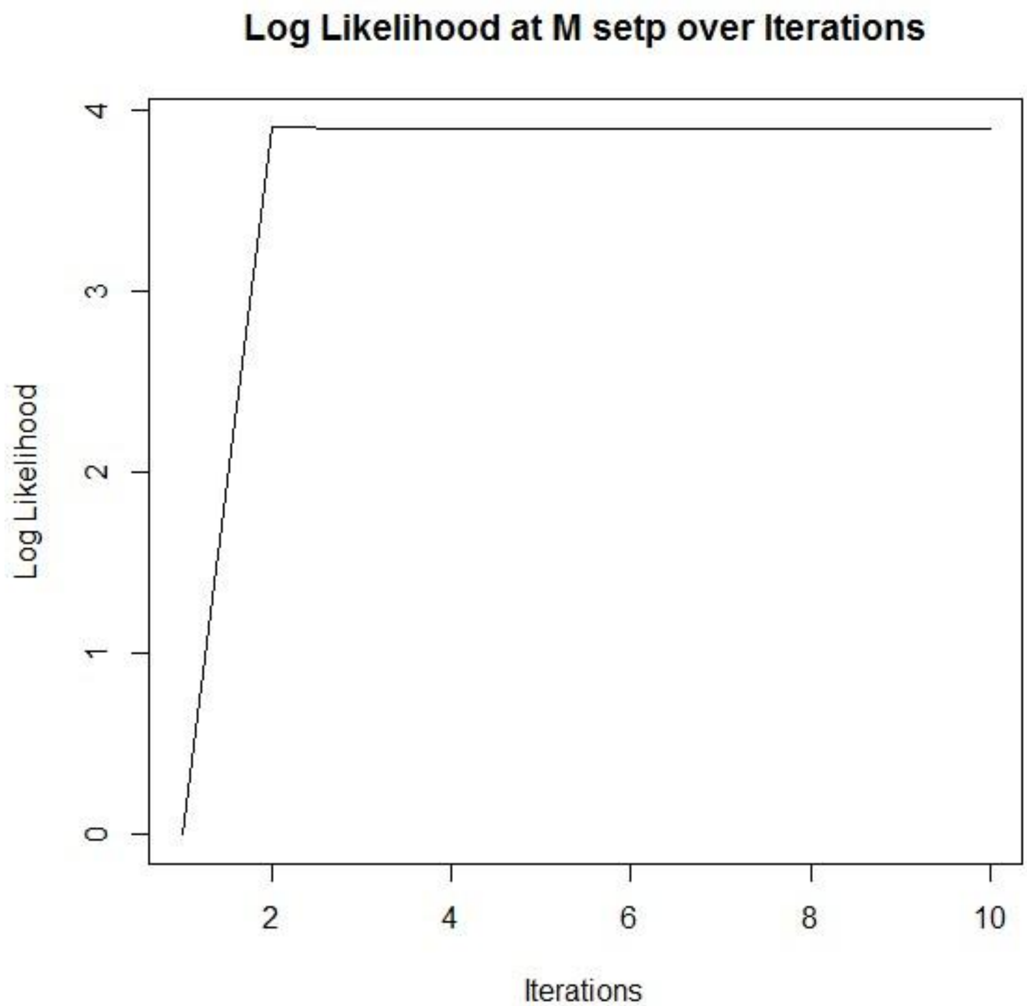

Figure 2.7: The change of log likelihood at M step over iterations for the mixture of two.

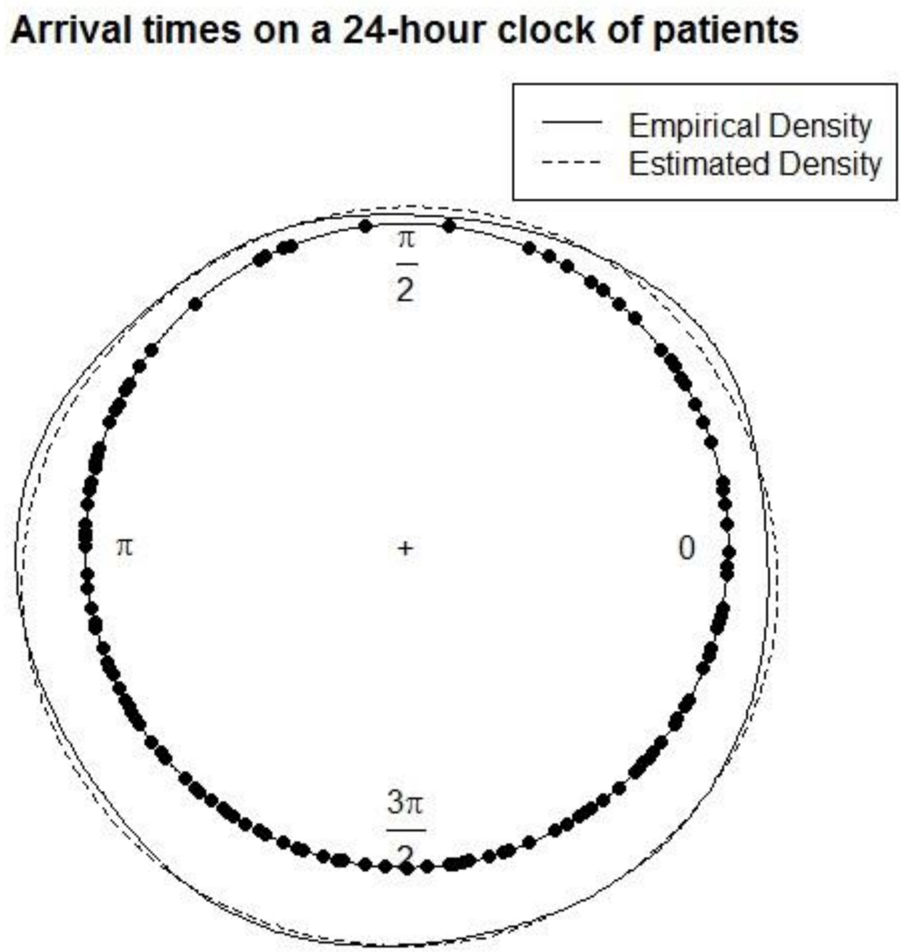

Figure 2.8: Comparison of the empirical density from the data and the estimated density from the mixture of two.

Thirdly, I fit the data set with a mixture of three von Mises distributions. The estimated parameters are summarized in Table 2.5. Figure 2.9 shows the log likelihood increases over the iterations and reaches the maximum at the 13^{th} iteration, indicating the convergence. The model suggests that the first cluster is at 0.6210, which is about 2:15 a.m.. This is a small cluster, where the proportion is only 8.76%. However, this is a relatively tight cluster because the concentration parameter is relatively large, namely 18.41. The second cluster is at 3.3085, which is about 12:40 p.m. This is a median cluster, where the proportion is 34.9%. The third cluster is at 5.0676, which is about 19:15 p.m.. This is the largest cluster among all three, where the proportion is

56.35%. Both the second and the third cluster are relatively sparse, where the concentration parameters are 2.23 and 1.57, respectively. Figure 2.10 shows the comparison of the empirical density from the data and the estimated density from the mixture of three von Mises distributions.

In order to perform the goodness of fit test, 500 data are simulated from the mixture of three von Mises distributions with the estimated parameters shown in Table 2.5. The simulated data are compared with the observed data using Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic is obtained to be 0.0597 and the p-value is 0.5846. Therefore, I fail to reject the null hypothesis and conclude that two sets of data are from the same distribution. So the mixture of three von Mises distributions gives good estimates fitting the data well.

Parameters	μ_1	μ_2	μ_3	λ_1	λ_2	λ_3	p_1	p_2	p_3
Estimations	0.6210	3.3085	5.0676	18.41	2.23	1.57	0.0876	0.3490	0.5635

Table 2.5: The summary of estimated parameters of a mixture of three von Mises distributions

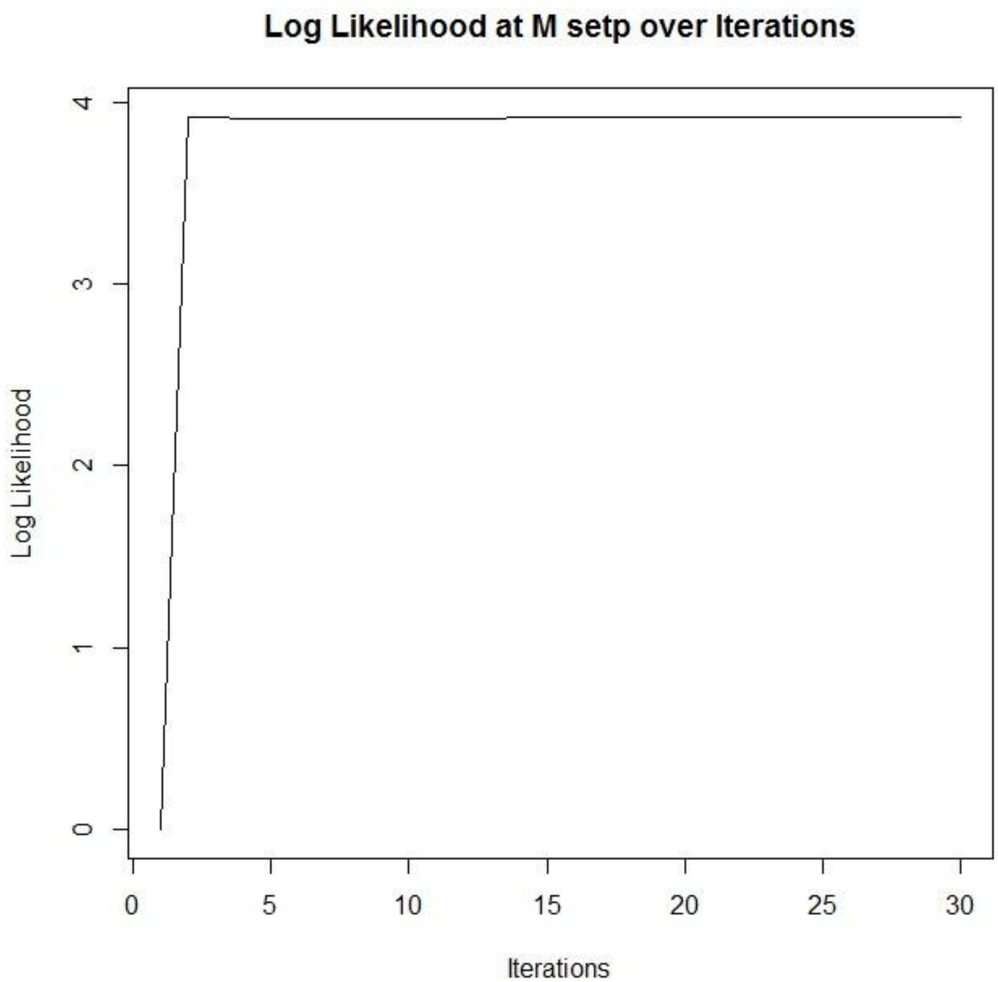

Figure 2.9: The change of log likelihood at M step over iterations for the mixture of three.

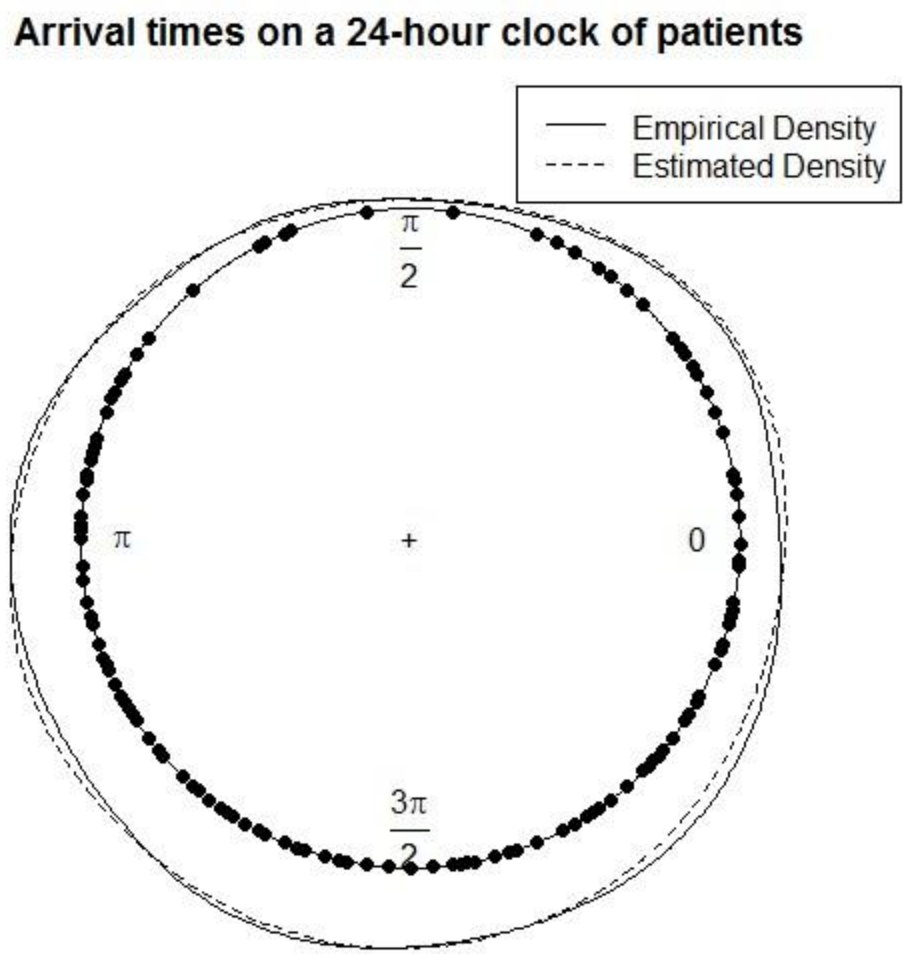

Figure 2.10: Comparison of the empirical density from the data and the estimated density from the mixture of three.

At last, I fit the data set with a mixture of four von Mises distributions. The estimated parameters are summarized in Table 2.6. Figure 2.11 shows the log likelihood increases over the iterations and reaches the maximum at the 18th iteration, indicating the convergence. The model suggests that the first cluster is at 0.6262, which is about 2:20 a.m.. This is a small cluster, where the proportion is only 11.74%. However, this is a relatively tight cluster because the concentration parameter is relatively large, namely 13.06. The second cluster is at 3.0778, which is about 11:45 a.m. The third cluster is at 4.3209, which is about 16:30 p.m.. And the fourth cluster is at 5.5184, which is about 21:00 p.m.. The second, the third, and the fourth cluster are relatively sparse, where the concentration parameters are 2.78, 4.66, and 3.84, respectively.

These three clusters are relatively in the same size, where the proportions are 31.33%, 27.01%, and 29.93%, respectively. Figure 2.12 shows the comparison of the empirical density from the data and the estimated density from the mixture of four von Mises distributions.

Again, in order to perform the goodness of fit test, 500 data are simulated from the mixture of four von Mises distributions with the estimated parameters shown in Table 2.6. The simulated data are compared with the observed data using Kolmogorov-Smirnov test. The Kolmogorov-Smirnov statistic is obtained to be 0.0558 and the p-value is 0.6698. Therefore, I fail to reject the null hypothesis and conclude that two sets of data are from the same distribution. So the mixture of four von Mises distributions gives good estimates fitting the data well.

Parameters	μ_1	μ_2	μ_3	μ_4	λ_1	λ_2	λ_3	λ_4
Estimations	0.6262	3.0778	4.3209	5.5184	13.06	2.78	4.66	3.84
Parameters	p_1	p_2	p_3	p_4				
Estimations	0.1174	0.3133	0.2701	0.2993				

Table 2.6: The summary of estimated parameters of a mixture of four von Mises distributions

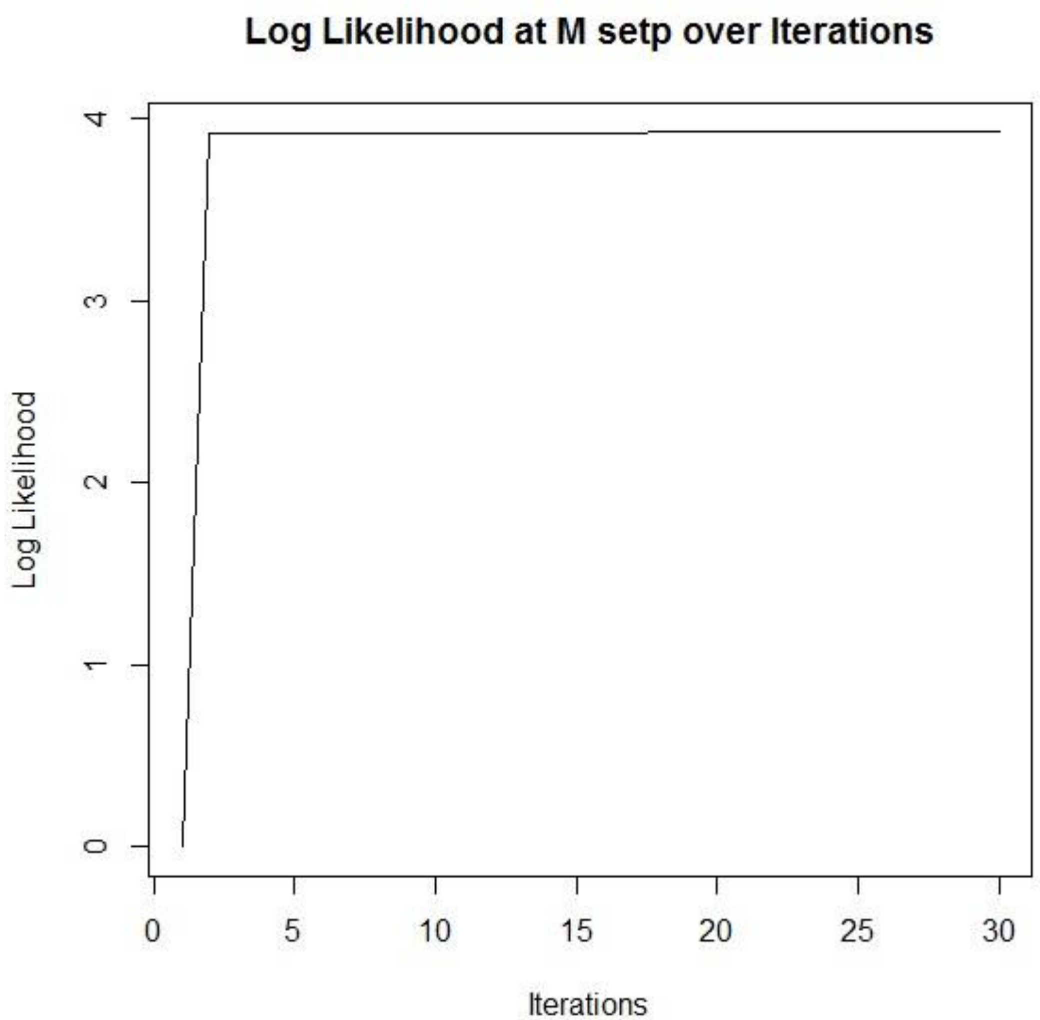

Figure 2.11: The change of log likelihood at M step over iterations for the mixture of four.

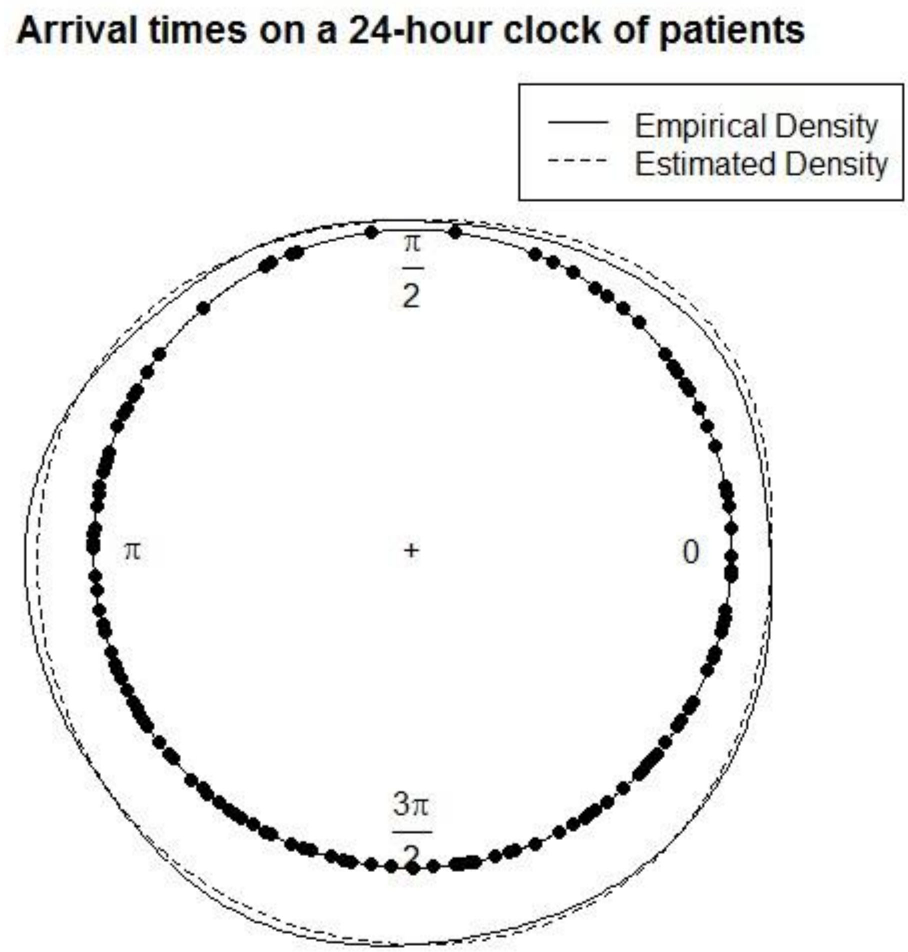

Figure 2.12: Comparison of the empirical density from the data and the estimated density from the mixture of four.

All four models demonstrate a good fit to the data set because I fail to reject the null hypothesis when I perform Kolmogorov-Smirnov tests for all four models. In order to find the best model among these four, Bayesian Information Criterion (BIC) test is used. Table 2.7 summarizes the BIC value obtained from each model. Among these four models, the mixture of two von Mises distributions has the highest BIC value, which is 511.2, indicating it is the worst fit. When fitting the data with either one von Mises distributions or a mixture of three von Mises distributions, the BIC values are very closed to each other. According to Figure 2.6 and Figure 2.10, although the estimated density from the mixture of three von Mises distributions seems to match the empirical density better, introducing more parameters in the mixture of three von Mises distributions outcompetes the decrease in the summation of logarithm of the likelihoods of the parameters given the data set, thus it results similar BIC values from these two models. The BIC value obtained from the mixture of four von Mises distributions is the smallest among four, which is 445.5254, thus indicating the mixture of four von Mises distribution gives the best fit to this data set.

	MLE	**Mixture of two**	**Mixture of three**	**Mixture of four**
BIC	475.9472	511.2	475.2903	445.5254

Table 2.7: The summary of BIC value obtained from each model

Chapter 3: Particle Filter Estimation for the Dynamic Model of Circular Data

INTRODUCTION

Over the years, the state-space models have been considered and implemented for modeling observations made over time. Consider the standard Markovian state-space model (West and Harrison 1997)

$$x_{t+1} \sim f(x_{t+1}|x_t) \qquad \text{(state evolution density)},$$

$$y_{t+1} \sim g(y_{t+1}|x_{t+1}) \qquad \text{(observation density)},$$

where $\{x_t\}$ are unobserved states (latent variables) of the system and $\{y_t\}$ are observations made over some time interval $t \in \{1,2,\cdots,T\}$. $f(\cdot\,|\,\cdot)$ and $g(\cdot\,|\,\cdot)$ are state evolution and observation densities that are pre-specified, which may be non-Gaussian and involve nonlinearity.

There are two major statistical inference problems associated with state space models, assuming the knowledge of all the parameters. The first concern is sequential estimation of the filtering distribution $p(x_t|y_{1:t})$, which is characterized by the joint posterior distribution of states at each point in time. In principle, updating the filtering density can be done using the standard filtering recursions. We first update $p(x_t|y_{1:t-1}) = \int p(x_{t-1}|y_{1:t-1})\, f(x_t|x_{t-1}) dx_{t-1}$, and then update $p(x_t|y_{1:t}) = p(x_t|y_{1:t-1}) g(y_t|x_t)/p(y_t|y_{1:t})$ .The second concern is state smoothing, which is characterized by the distribution of the states, conditional on all available data. Similarly, smoothing can be performed recursively backward in time using the smoothing formula $p(x_t|y_{1:T}) = \int p(x_{t+1}|y_{1:T})\, p(x_t|y_{1:t-1}) g(y_t|x_t)/p(y_t|y_{1:t})\, dx_{t+1}$ (Godsill et. al 2004; Carvalho et. al 2010).

Extensive research has been done over the past decade about the inference in general state-space models. Monte Carlo strategies based on sequential importance sampling, known generically as particle filters, have been emerging rapidly because these methods have been proven for many examples, including highly nonlinear models that are not easily implemented using standard Markov chain Monte Carlo. In particle filtering methods, the filtering density can be approximated with an empirical distribution formed from points masses, or particles. These particles at time t can be updated efficiently to particles at time t+1 using importance sampling and re-sampling methods.

In this chapter, I present a situation where circular data are observed from the Markovian state-space model. Both sequential estimation of the filtering distribution and backward smoothing using particle filter methods are reviewed and implemented for the circular data.

FORWARD FILTERING BACKWARD SMOOTHING (FFBS)

Consider a set of circular observations $y_t, t = 1, \cdots, T$, following a first order dynamic linear model driven by a discrete latent state θ_t. In particular, we have

$$y_{t+1} = \theta_{t+1} + \nu, \nu \sim vM(0, \lambda_1),$$
$$\theta_{t+1} = \alpha + \beta\theta_t + \varepsilon, \varepsilon \sim vM(0, \lambda_2),$$

where α is a circular measure and β is a constant. The two noise terms are assumed to follow two von Mises distributions whose center directions are 0 and the concentration parameters are λ_1 and λ_2, respectively. λ_1 and λ_2 are assumed to be independent.

By observing the circular data at each time point, assuming the knowledge of all the parameters, the purpose is to first sequentially estimate the filtering distribution $p(\theta_t|y_{1:t})$ at each time point. Then after one sequential pass through the data, we obtain the samples $\{\theta_t{}^{(i)}\}_{i=1}^{M}$ at each time point, where $t = 1, \cdots, T$. Secondly, conditional on all available information, a direct backward sequential pass is used to obtain full smoothing distributions. In particular, for each time point $t = T - 1, \cdots, 1$, resample θ_t from $\{\theta_t{}^{(i)}\}_{i=1}^{M}$ with weights $\pi_{t-1|t}^{(i)} \propto p\left(\theta_{t-1}^{(i)}|\theta_t\right) \sim vM\left(\theta_t{}^{(i)} - \alpha, \lambda_2\right)$. After one backwards smoothing, we obtain all the latent variables $\theta_{T:1}$.

Algorithm for FFBS

Step 0 (Initiation). Choose θ_0 arbitrarily.

Step 1 (Sample). Draw $\{\theta_{t-1}{}^{(i)}\}_{i=1}^{M}$ from vM$(\theta_{t-1}, \lambda_1)$ at time t-1.

Step 2 (Propagate). $\{\theta_{t-1}{}^{(i)}\}_{i=1}^{M}$ to $\left\{\tilde{\theta}_t{}^{(i)}\right\}_{i=1}^{M}$ via $\tilde{\theta}_t{}^{(i)} \sim vM(\beta\theta_{t-1} + \alpha, \lambda_2)$. Here $\{\theta_{t-1}{}^{(i)}\}_{i=1}^{M}$ is approximately $p(\theta_t|\alpha, \beta, \lambda_1, \lambda_2, y_{1:t-1})$.

Step 3 (Resample). $\left\{\theta_t^{(i)}\right\}_{i=1}^{M}$ to $\left\{\tilde{\theta}_t^{(i)}\right\}_{i=1}^{M}$ with weights $\omega_t^{(i)} \propto p\left(y_t|\tilde{\theta}_t^{(i)}\right) \sim vM\left(\tilde{\theta}_t^{(i)}, \lambda_1\right)$. Here $\left\{\theta_t^{(i)}\right\}_{i=1}^{M}$ is approximately $p(\theta_t|\alpha, \beta, \lambda_1, \lambda_2, y_{1:t})$.

Step 4 (Backward Smoothing). θ_t to θ_{t-1} from $\left\{\theta_{t-1}^{(i)}\right\}_{i=1}^{M}$ with weights $\pi_{t-1|t}^{(i)} \propto p\left(\theta_{t-1}^{(i)}|\theta_t\right) \sim vM\left(\theta_t^{(i)} - \alpha \big/ \beta, \lambda_2\right)$.

EXAMPLE WITH SIMULATED DATA

Assume α is $\frac{\pi}{15}$, which is 12 degrees. β is 0.8. λ_1 is 20 and λ_2 is 10. Let the initial state θ_1 be $\frac{\pi}{2}$. $\theta_{2:T}$ are simulated via the conditional distribution $p(\theta_t|\theta_{t-1}) \sim vM\left(0.8 * \theta_{t-1} + \frac{\pi}{15}, 10\right)$, where T=300. Then the observed data $y_{1:T}$ are simulated via the conditional distribution $p(y_t|\theta_t) \sim vM(\theta_t, 20)$. Based on the simulated observed data $y_{1:T}$, latent variables, $\theta_{1:T}$, are estimated through FFBS method.

In order to assess the performance of FFBS algorithm, the estimated latent variables are compared with the simulated ones at each time point. Because the latent variables are circular data, arc distance is used to summarize the difference between each pair at each time point. Arc distance is defined as the shortest distance between two points along the circle making up the arc. The maximum arc distance between any two points on a unit circle is π.

Figure 3.1 shows the arc distances of estimated latent variables and simulated ones at each time point after one sequential filter through the data. The solid red line, which is 0.212 and about 12.147 degrees, represents the mean arc distance for all 300 time points. It means that on average, the estimated latent variables and simulated ones are 0.212 or 12.147 degrees apart from each other. Table 3.1 shows the summary of these 300 arc distances and Figure 3.2 is the histogram of these 300 arc distances with the mean as the solid red line.

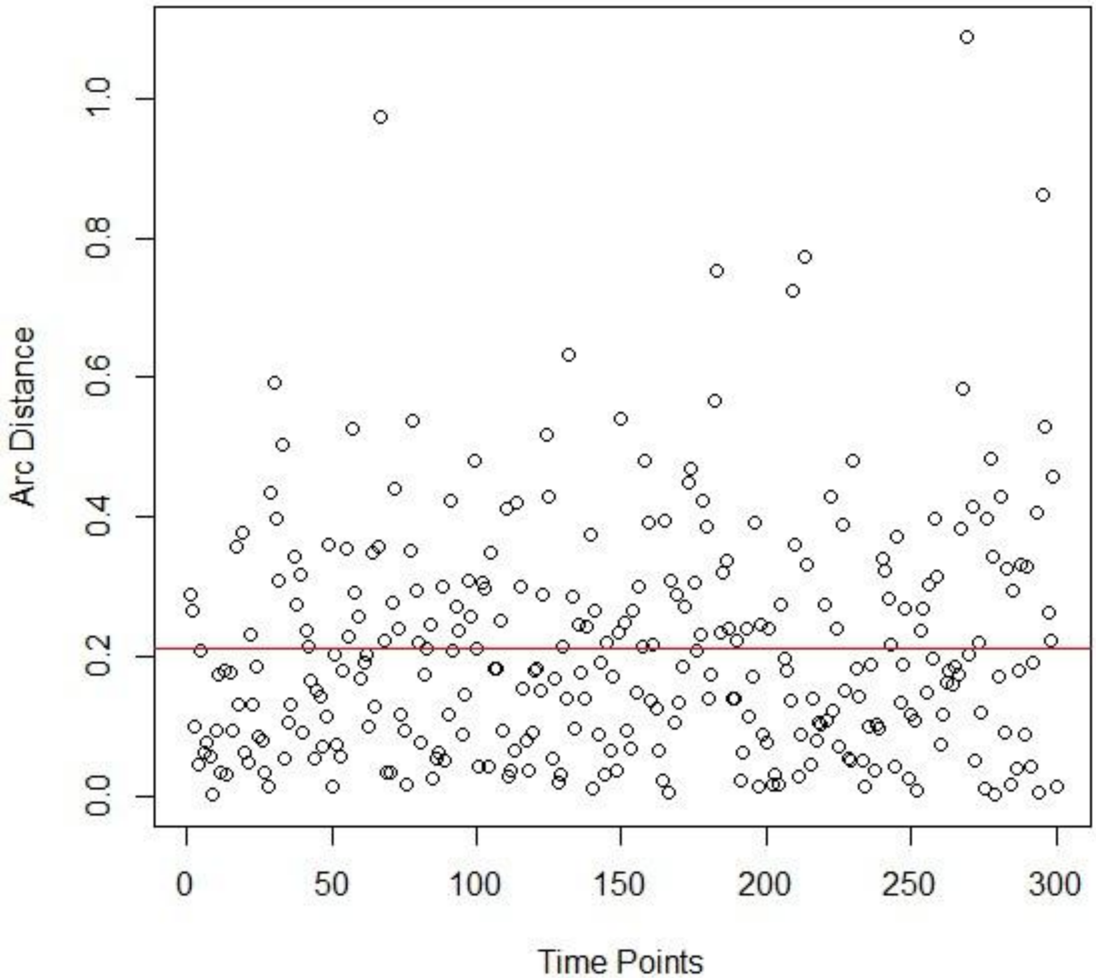

Figure 3.1: The plot of arc distances between estimated latent variables and simulated ones at each time point after one sequential filter through the data

Minimum	1st Quantile	Median	Mean	3rd Quantile	Maximum
0.0005058	0.08762	0.1808	0.212	0.2965	1.087

Table 3.1: The summary of arc distances of 300 time points after one sequential filter through the data

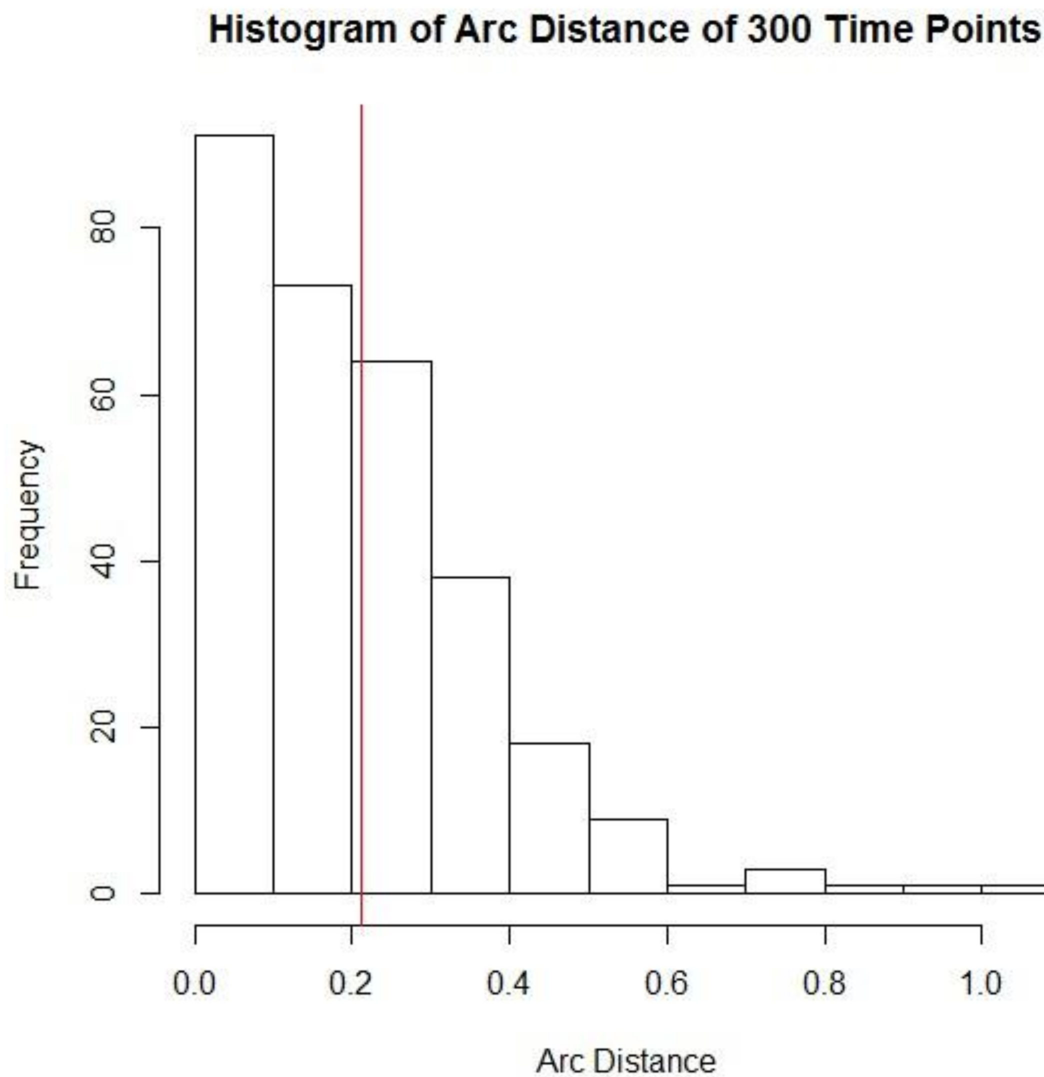

Figure 3.2: The histogram of arc distances between estimated latent variables and simulated ones at each time point after one sequential filter through the data

Figure 3.3 shows the arc distances of estimated latent variables and simulated ones at each time point after a direct backward sequential smoothing is used. The solid red line, which is 0.2015 and about 11.5 degrees, represents the mean arc distance for all 300 time points. It means that on average, the estimated latent variables and simulated ones are 0.2015 or 11.5 degrees apart from each other. Table 3.2 shows the summary of these 300 arc distances and Figure 3.4 is the histogram of these 300 arc distances with the mean as the solid red line.

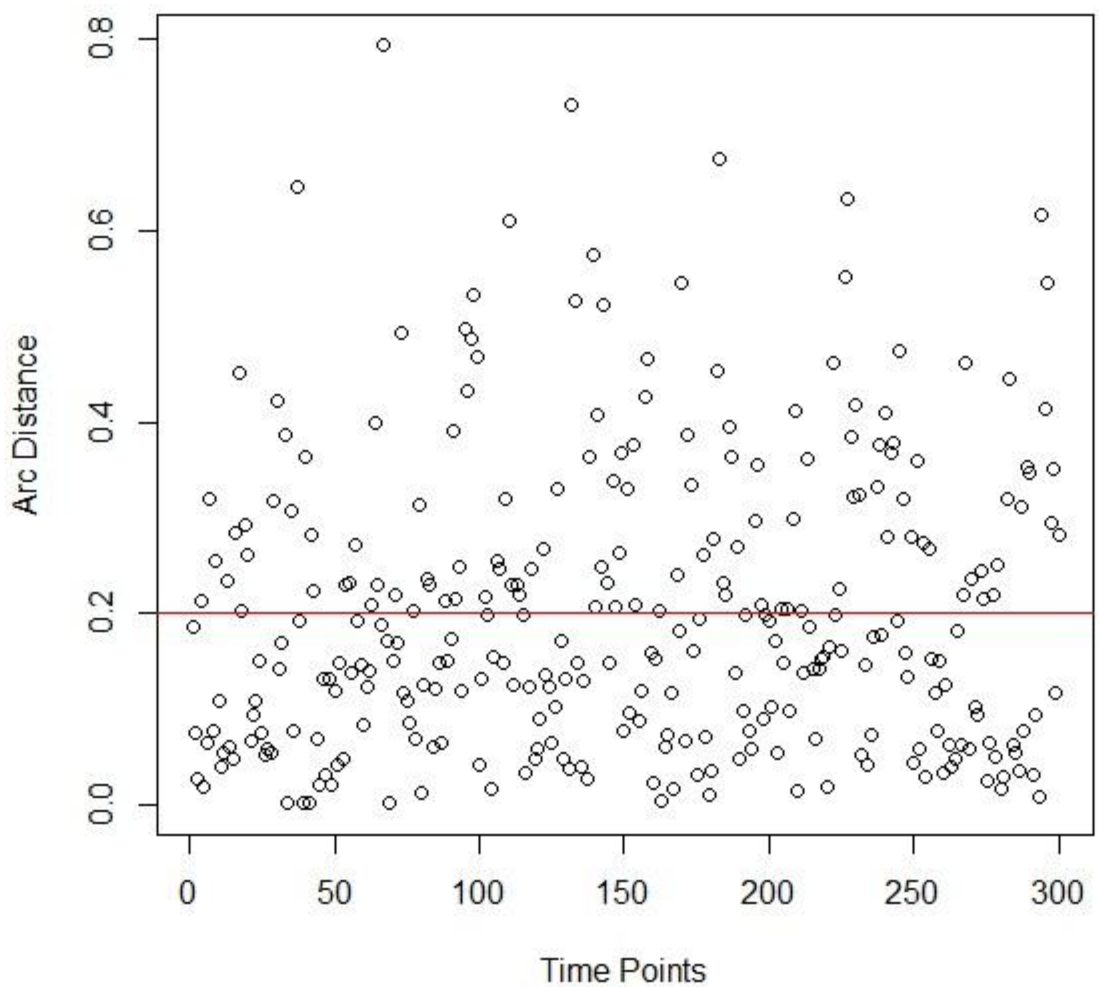

Figure 3.3: The plot of arc distances between estimated latent variables and simulated ones at each time point after FFBS.

Minimum	**1st Quantile**	**Median**	**Mean**	**3rd Quantile**	**Maximum**
0.001055	0.07667	0.1703	0.2015	0.2804	0.7935

Table 3.2: The summary of arc distances of 300 time points after FFBS.

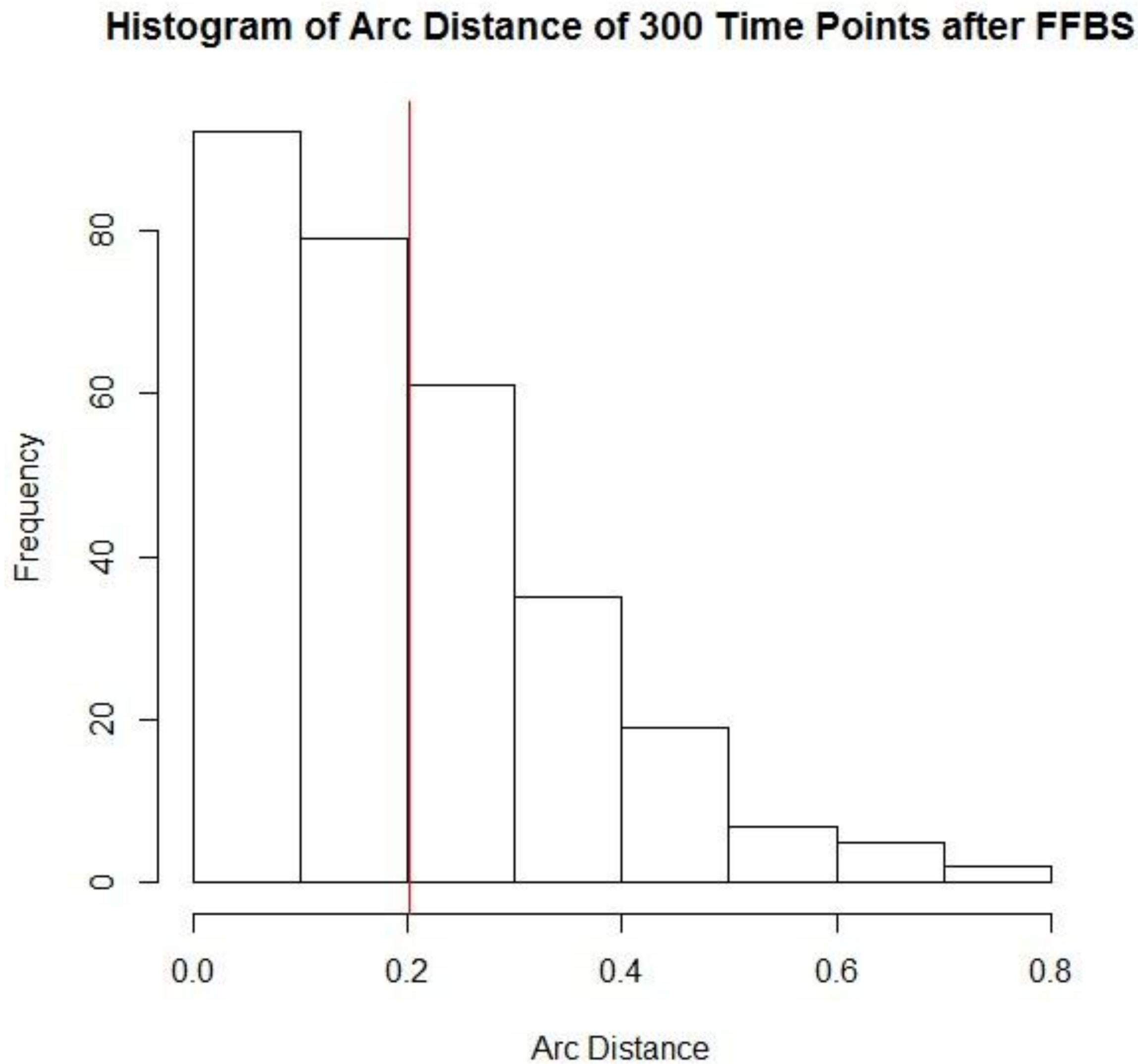

Figure 3.4: The histogram of arc distances between estimated latent variables and simulated ones at each time point after FFBS.

In order to assess the performances of only one sequential filter through the data vs. additional backward sequential smoothing, Figure 3.2 and Figure 3.4 are compared. As shown in Figure 3.4, there is no arc distance greater than 0.8 if FFBS is used. The arc distance is much more skewed toward the left in Figure 3.4, indicating much smaller differences between estimated latent variables and the true simulated ones. The differences in Figure 3.4 and Figure 3.2 can demonstrate the improvements in estimations of latent variables when FFBS, particularly backward smoothing through the data, is used.

References

Akesson, S., Karlsson L., Walinder, G., and T. Alerstam. 1996. Bimodal orientation and the occurrence of temporary reverse bird migration during autumn in south Scandinavia. Behav Ecol Soclobiol 38: 293-302.

Calderara S., Prati A., and R. Cucchiara. 2011. Mixtures of von Mises distributions for people trajectory shape analysis. IEEE Transactions on Circuits and Systems for Video Technology 21(4): 457-471.

Capaccioni B., Valentini L., Rocchi M.B.L., Nappi G., and D. Sarocchi. 1997. Image analysis and circular statistics for shape-fabric analysis: applications to lithified ignimbrites. Bull Volcanol 58: 501-514.

Carvalho C.M., Johannes M.S., Lopes H.F., and N.G. Polson. 2010. Particle learning and smoothing. Statistical Science 25(1): 88-106.

Damien P. and S. Walker. 1999. A full Bayesian analysis of circular data using the von Mises distribution. Canadian Journal of Statistics 27: 291-298.

Doucet, A., de Freitas, J., and N. Gordon. 2001. Sequential Monte Carlo methods in practice. Springer, New York. MR 1847783.

Fisher, N.I. 1993. Statistical analysis of circular data. Cambridge University Press.

Godsill S.J., Doucet A. and M. West. 2004. Monte Carlo Smoothing for Nonlinear Time Series. Journal of the American Statistical Association 99(465): 156-168.

West M. and J. Harrison. 1997. Bayesian Forecasting and Dynamic Models (2nd ed.). New York: Springer-Verlag.